



## Mots. Les langages du politique

102 | 2013

Les discours sur l'enseignement supérieur et la recherche

---

### *Excellence. Une analyse logométrique*

Matthias Tauveron

---



#### Édition électronique

URL : <http://journals.openedition.org/mots/21387>

DOI : 10.4000/mots.21387

ISSN : 1960-6001

#### Éditeur

ENS Éditions

#### Édition imprimée

Date de publication : 2 septembre 2013

ISBN : 978-2-84788-414-2

ISSN : 0243-6450

#### Référence électronique

Matthias Tauveron, « *Excellence. Une analyse logométrique* », *Mots. Les langages du politique* [En ligne], 102 | 2013, mis en ligne le 02 septembre 2015, consulté le 20 avril 2019. URL : <http://journals.openedition.org/mots/21387> ; DOI : 10.4000/mots.21387

---

**Matthias Tauveron**

## ***Excellence.* Une analyse logométrique**

La communication publique concernant l'Université – notamment celle venant des institutions officielles – joue toujours un jeu dangereux, entre le choix d'une communication strictement technique, informative, et celui de promouvoir un établissement, un système, une gestion, afin de le doter d'une visibilité (Granget, 2009). Dans ce jeu, le nom *excellence* prend une place envahissante, notamment depuis 2009 dans les discours sur l'enseignement supérieur français. *Excellence* contribue en effet depuis cette date à former les expressions officielles *Laboratoire d'excellence*, *Initiative d'excellence* et *Équipement d'excellence*. Pourtant, dans les discours mêmes où ces terminologies sont utilisées, ce nom garde encore son sens non technique pour faire référence à un niveau de compétence maximal, idéal.

### **La rhétorique de l'excellence dans les discours sur l'enseignement supérieur**

L'usage de ce nom dans ces discours suscite pour nous deux grandes familles de questions. La première est liée à son ambivalence, issue de son appartenance à différents domaines : comment voir, comment montrer qu'il relève pleinement à la fois de la sphère technique (nouvelle gestion de la recherche et de l'Université, questions administratives, financières) et d'une sphère « idéale » attachée à des ambitions de visibilité, de réussite, c'est-à-dire à des idéaux dégagés des contraintes matérielles ?

La seconde est celle de son inscription dans les discours. Nous avons choisi de nous focaliser sur un texte officiel – l'appel à projet de 2011 pour les initiatives d'excellence – qui montre à la fois une dimension technique (précisant les modalités de rédaction de projets pour les équipes candidates aux financements en question) et une dimension argumentative, mettant en place l'idéal de la recherche qu'il appelle *excellence*. Il faut souligner que sa dimension

argumentative tombe dans le même écueil que le discours expert tel que décrit par Cussó et Gobin (2008, p. 7). En effet, sans fournir d'idée qui puisse faire polémique, il se contente de mettre en avant des idées qui ne sauraient essayer de contre-argumentation. Qui serait contre l'excellence ? Ce discours a donc a priori la même ambivalence que le nom *excellence*, étant à la fois technique et non technique, procédant par sa seule autorité pour appuyer l'idée qu'il prône. L'inscription du nom *excellence* dans un tel discours pose alors la question suivante : quelle est la contribution de ce nom au sens textuel global ? Comment ce nom s'intègre-t-il à la textualité de ce document pour former un ensemble ?

Cette étude de l'intégration d'un mot dans un discours passe, selon nous, à la fois par l'étude des occurrences – relevés et statistiques concernant la présence de mots dans les textes – et par celle des cooccurrences. En l'espèce, l'étude de la « cooccurrence généralisée » (Viprey, 2006), qui synthétise l'ensemble des relations de cooccurrence entre les mots d'un texte donné, montre la manière dont le lexique est utilisé dans un texte (voir figure 2 en annexe)<sup>1</sup>. Tentant de faire un pont entre la sémantique (locale) du mot et la sémantique (globale) du texte, elle montre comment la représentation de l'excellence peut être plurielle au sein d'un même discours (voir Rémi-Giraud, 2010). En étudiant l'intrication du nom *excellence* dans le réseau des mots du texte, on peut s'interroger sur l'éventuelle « *autonomie sémantique* » (Gobin, Deroubaix, 2010) du concept sous-jacent, et voir ainsi si et comment il se voit investi par les dimensions techniques, institutionnelles ou humaines présentes dans le même discours.

Si nous étudions la cooccurrence généralisée telle que décrite par Viprey (2006), nous préférons à sa représentation le *graphe de cooccurrence* (voir Véronis, 2004, et figure 2 en annexe). Grâce à cet outil, nous pouvons notamment montrer que le lexique d'un texte a une organisation *modulaire* (voir *infra*, « Analyse lexicométrique... »). C'est-à-dire que le lexique – et donc le contenu du texte – peut être divisé de manière automatique en un nombre plus ou moins grand d'ensembles (dits « communautés »), chacune représentée par un niveau de gris sur les illustrations fournies ici. Suivant des modalités que nous dévoilerons plus loin (même section, *infra*), l'utilisateur a la possibilité de choisir une plus ou moins grande finesse pour ce découpage.

Pour la présente analyse, nous proposons une nouvelle interprétation de cette modularité, que nous appellerons le *feuilletage*. On entendra par là que le lexique du texte peut être considéré comme une superposition de couches d'ensembles de mots. La figure 1 (en annexe) montre la descente, depuis une

1. Parmi toutes les manières d'envisager la cooccurrence qui ont été proposées, nous nous focalisons sur celle-ci pour deux raisons que nous reprenons à l'auteur. La première est que la cooccurrence généralisée est un objet que l'on peut étudier qualitativement, et non seulement quantitativement. La seconde est qu'elle permet d'aborder ce qui est particulier à un texte donné (son vocabulaire), en restant autonome par rapport à ce que propose la langue.

couche superficielle à une couche d'analyse plus profonde de l'organisation du lexique d'un même texte, que permet cette analyse. L'esprit de cette descente d'une couche à l'autre est illustré sur le tableau 1 :



Tableau 1. Représentation du feuilletage lexical d'un texte

### Analyse lexicométrique par le graphe de cooccurrence

Le graphe de cooccurrence est construit par un programme Perl écrit par nos soins qui parcourt l'ensemble du texte à l'aide d'une fenêtre paramétrable circonscrivant les relations de cooccurrence<sup>2</sup>. Nous optons ici pour une relation de cooccurrence symétrique (malgré Luong *et al.*, 2010), et telle que, si deux noms sont présents dans la même phrase ou dans deux phrases voisines, ils sont cooccurents<sup>3</sup>. Notre programme synthétise ensuite ces résultats dans deux fichiers au format CSV utilisables par Gephi<sup>4</sup>. Une représentation du graphe de cooccurrence montre alors l'ensemble des mots choisis (dans notre cas, les noms communs du texte), chacun sous la forme d'un « nœud » (cercle muni d'une étiquette). Deux nœuds sont reliés par un « lien » (représenté par un trait) si les deux mots en question sont cooccurents. La taille des cercles et la largeur des liens représentent l'importance, respectivement, des nœuds et des liens<sup>5</sup>.

2. Le texte utilisé est lemmatisé, ce qui se justifie assez bien par le fait que nous cherchons à identifier les contenus d'ensemble du texte, ce qui, selon nous, fonctionne mieux en laissant de côté la variation morphosyntaxique. Pour des raisons de lisibilité, nous suivons la recommandation d'Étienne Brunet (communication personnelle, mai 2012) de nous cantonner aux seuls substantifs. Des problèmes inhérents à notre lemmatiseur font apparaître par erreur des adjectifs dans nos graphes (ainsi *définitif* et *sélectionné* sur la figure 3 en annexe). Nous nous efforçons de ne pas tenir compte de ces formes dans nos analyses.
3. Le nombre de relations de cooccurrence dans un texte étant en général très élevé, nous avons élagué le graphe en deux étapes. En premier lieu, ne sont considérés que les lemmes apparaissant au moins 20 fois dans le texte. Puis ne sont retenus que les liens de cooccurrence attestés au moins 50 fois.
4. Logiciel libre et gratuit de traitement de graphes ([www.gephi.org](http://www.gephi.org)). Nous utilisons pour la présente étude la version 0.8.1-beta, première version à implémenter la gestion de la modularité telle que nous l'utilisons ici.
5. Dans cette étude, l'importance des nœuds est mesurée par la *betweenness centrality*, qui est plus élevée à mesure que le nœud en question est un « carrefour » du graphe. C'est-à-dire que plus le nœud en question apparaît sur le plus court chemin reliant deux nœuds quelconques du graphe, plus cette centralité est élevée. Dans le cas des graphes de cooccurrence, cette propriété est assez bien corrélée à la fréquence (Tauveron, 2012). L'importance d'un lien est mesurée directement : elle est proportionnelle à la fréquence du lien de cooccurrence qu'il représente.

L'identification des couches de lexique dans le texte passe par l'examen de la modularité du graphe de cooccurrence, c'est-à-dire la détection de communautés de mots dans ce graphe. Déceler de telles communautés dans un graphe consiste à diviser un graphe en différentes sous-parties de façon pertinente au regard de la constitution du graphe. C'est-à-dire qu'il ne s'agit pas de découper arbitrairement le graphe en autant de parties que le souhaite l'analyste (Newman, 2006) : le but est de déceler les articulations pour ainsi dire « naturelles » de l'objet (Blondel *et al.*, 2008), et ainsi trouver les sous-parties qui le constituent spontanément. Ces articulations naturelles sont le résultat de la disposition des liens entre les nœuds. Chaque communauté de nœuds est en effet caractérisée par le fait que les nœuds qui la constituent ont plus de liens entre eux qu'ils n'en ont avec les nœuds extérieurs à la communauté (Chen, Ji, 2010)<sup>6</sup>.

S'agissant du graphe de cooccurrence qui nous intéresse ici, une communauté de mots est donc un ensemble de mots qui ont tendance à être cooccurents préférentiellement les uns avec les autres deux à deux, et à être moins cooccurents avec d'autres mots appartenant à d'autres communautés. Ainsi, cette méthode permet de découper le lexique d'un texte donné en différents sous-ensembles, de façon pertinente au regard de l'usage des mots dans ce texte. Elle montre donc – sans intervention humaine, en tout cas sans interprétation a priori – comment le lexique du texte se répartit en différentes familles, et ceci sans faire d'hypothèse à partir de connaissances sur la langue, le discours ou le genre.

Gephi utilise l'algorithme proposé par Blondel *et al.* (2008), dit parfois *algorithme de Louvain*. Cet algorithme a un fonctionnement récursif (Blondel, 2011) qui illustre bien le caractère à la fois local et global de la modularité. Dans un premier temps, il construit en effet de petites communautés faites d'agglomération de nœuds voisins (part locale). Puis il monte peu à peu dans la hiérarchie en faisant fusionner entre elles des communautés voisines et bien reliées (part globale).

Comme le laisse supposer cette description, l'algorithme de Louvain n'est pas configurable et donne, pour un graphe donné, un unique résultat. Lambiotte *et al.* (2009) en ont proposé un raffinement, qui permet de paramétrer celui-ci et de régler grossièrement la taille des communautés obtenues. C'est au prix de ce paramétrage que l'on peut – selon nous – entrer dans le feuilletage de la cooccurrence généralisée.

Le point de départ de l'algorithme de Lambiotte *et al.* est tout à fait différent de celui de l'algorithme de Louvain. Il repose sur la métaphore du marcheur

6. Le fait définitoire techniquement essentiel est qu'un tel découpage est une *partition* du graphe, c'est-à-dire que chaque nœud appartient à exactement une communauté, sans que celles-ci ne se chevauchent ou que le découpage ne laisse de restes.

aléatoire. L'algorithme place au départ un personnage sur un nœud  $N$  donné du graphe, et laisse celui-ci marcher en suivant les liens du graphe. À chaque nœud rencontré, le marcheur décide aléatoirement lequel des liens afférant à ce nœud il va emprunter. Après un certain temps ajustable par l'utilisateur, on arrête le marcheur. L'ensemble des nœuds qu'il a visités forme, par définition, la communauté dont fait partie le nœud  $N$ . L'opération est recommencée autant de fois que nécessaire en prenant d'autres nœuds de départ jusqu'à former une partition du graphe en communautés. Pour des raisons qui apparaîtront *infra*, le temps de parcours laissé au marcheur est appelé *résolution* de l'algorithme.

La démonstration que fournissent Lambiotte *et al.* montre que leur algorithme, dans les résultats qu'il fournit, est interprétable comme un raffinement du travail de Blondel *et al.* (2008). Ils montrent en effet que, si on utilise leur algorithme avec une résolution valant 1.0, le résultat est exactement le même que l'algorithme de Louvain. L'intérêt repose en fait sur le choix de la valeur donnée à ce paramètre. Baisser la valeur de la résolution donne moins de temps au marcheur pour parcourir le graphe : il rencontrera ainsi moins de nœuds à chacun de ses parcours, et chaque communauté sera donc plus petite. Au contraire, augmenter cette valeur lui laisse plus de temps, le laisse parcourir plus de chemin à chaque fois, et forme donc des communautés plus grandes. Comme l'ensemble des communautés forme, par définition, une partition (voir note 6 *supra*), ces contraintes de taille ont pour contrepartie des contraintes de nombre :

- une faible valeur de la résolution donne des communautés plus petites, et donc plus nombreuses ;
- une forte valeur de la résolution fournit des communautés plus grandes, et donc moins nombreuses<sup>7</sup>.

Ainsi, c'est le choix de trois valeurs de résolution distinctes qui fournit trois représentations du même graphe de cooccurrence sur la figure 1 *supra*.

### *Proposition d'analyse : la superposition des couches de nœuds*

Nous proposons ici d'utiliser le raffinement décrit ci-dessus en multipliant les applications de cet algorithme sur le même graphe, à différentes valeurs de résolution. Ce faisant, on forme successivement différentes partitions du graphe, certaines faisant apparaître des communautés plus nombreuses (découpage plus fin), d'autres faisant apparaître des communautés moins nombreuses (découpage plus grossier). Chacun de ces réglages de résolution révèle pour nous une couche du graphe. Changer la valeur de ce paramètre per-

7. On remarquera donc qu'une valeur de résolution *plus élevée* conduit à une analyse *plus grossière* et, dans notre interprétation, à une *moins grande profondeur* de l'examen du graphe. Ce fait est tout à fait contre-intuitif et dommageable, mais il vient de la définition de la résolution, qui ne nous appartient pas.

met de changer la profondeur à laquelle on examine l'organisation du graphe, et donc l'organisation de la cooccurrence généralisée dans un texte donné.

Il faut signaler que, dans la grande majorité des cas, ces changements de résolution montrent une organisation hiérarchique de ces couches de communautés. C'est-à-dire (comme cela apparaît sur notre tableau 1) qu'une communauté à un niveau de profondeur donné va se séparer (ou non) en différentes communautés à un plus grand niveau de profondeur. Une communauté à un niveau inférieur a donc exactement une communauté « parent » au niveau supérieur. La situation différente, illustrée sur notre tableau 2 *infra*, dans laquelle les communautés d'un niveau inférieur peuvent avoir plusieurs « parents » au niveau supérieur, est exceptionnelle.



Tableau 2. Illustration d'une situation particulière dans le passage d'une couche à l'autre

## Plongée dans les couches du lexique d'un discours institutionnel

### *Corpus utilisé : thématique et rhétorique*

Le texte auquel est consacré la présente analyse est l'« Appel à projet pour les initiatives d'excellence » (désormais AAPIE), publié sur le site de l'AERES courant 2011<sup>8</sup>. Ce document détaille la procédure à suivre de la part des laboratoires candidats pour obtenir le statut de « pôle pluridisciplinaire d'excellence ». Ce texte a été choisi pour son utilisation du nom *excellence*, ainsi que pour son genre et sa rhétorique. En effet, étant un appel à projets, il a une certaine dimension injonctive (détaillant une procédure à suivre), et une dimension argumentative, valorisant le bien-fondé de cette nouvelle forme d'organisation et de financement de la recherche publique :

Une Initiative d'excellence assure ainsi la promotion et le développement d'un périmètre d'excellence et impulse autour de lui une dynamique de structuration du site par la mise en œuvre d'actions de recherche et de formation innovantes dans le cadre d'une gouvernance rénovée et performante. (p. 5)

Par ailleurs, et notre analyse cherchera à préciser ce point de vue, cette

8. En ligne [<http://www.agence-nationale-recherche.fr/investissementsdavenir/AAP-IDEX-2011.html>], consulté le 11 mars 2013. La clôture de l'appel à projet était fixée au 20 septembre 2011, ce qui permet de dater approximativement le document. Le texte contient 5 614 mots, dont 95 occurrences d'*excellence*.

double dimension se retrouve reflétée d'une certaine manière dans le lexique du texte. Ce texte fait en effet référence :

- d'une part, à une thématique textuelle, d'écriture et de lecture, en l'occurrence des projets rédigés par les candidats ; ce champ sémantique rejoint la dimension injonctive signalée plus haut ;
- d'autre part, à un ensemble d'idées touchant à l'organisation de la recherche, à l'objectif de celle-ci, dans une dimension à la fois technique et scientifique. Cette part du lexique est donc reliée étroitement à la dimension argumentative du texte.

Comme un certain nombre d'autres noms, *excellence* joue un rôle d'intermédiaire entre ces deux domaines. Il fait en effet partie de la terminologie institutionnelle (sous la forme des sigles tels que LabEx, IdEx, etc.) et, à ce titre, concerne directement la rédaction des projets, qui sont une des incarnations de cette dimension institutionnelle. Mais ceci ne retire en rien son côté idéologique, l'argumentation et la mise en place des idées dans ce texte officiel profitant de l'axiologie positive qu'il implique.

### *L'entrée dans les niveaux de profondeur*

La méthode proposée ici consiste à examiner le passage d'une couche de lexique à l'autre en relevant les changements dans l'organisation en communautés de ce lexique. Une des difficultés de réalisation est que rien ne permet de prévoir les valeurs adéquates de résolution à choisir pour obtenir un résultat intéressant<sup>9</sup>.

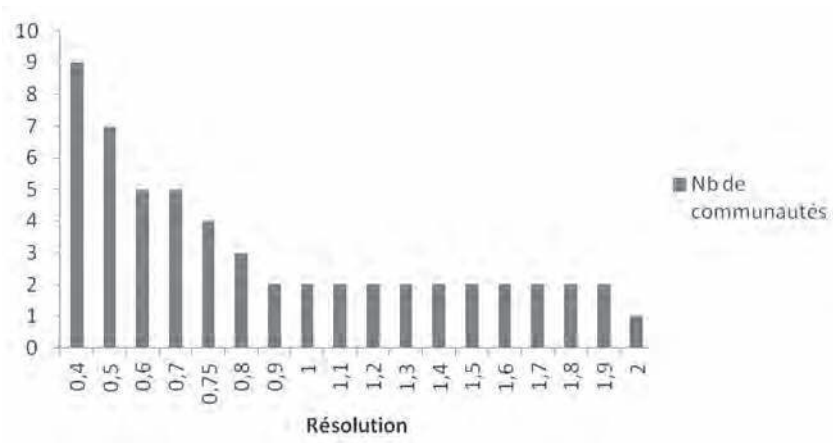
La méthode présentée plus haut a été appliquée sur le texte choisi, et donne le résultat présenté sur le graphique 1 *infra*. Comme on s'y attend au regard de la description de la résolution donnée plus haut, le nombre de communautés décroît à mesure que l'on augmente la résolution de l'algorithme. On peut donc passer de la couche extérieure de l'organisation du lexique du texte (avec une résolution de 2.0, qui donne une seule communauté), en passant par différentes couches successives, jusqu'à une résolution de 0.4, couche très profonde qui fait apparaître 9 communautés de mots.

Il faut attirer l'attention sur un fait supplémentaire : le découpage de graphe en deux communautés de nœuds est un découpage particulièrement stable, qui se maintient durablement lorsqu'on fait varier la résolution de 0.9 à 1.9. Cette grande stabilité témoigne sans doute de la pertinence de ce découpage au regard de la structure même du graphe. On rappelle en effet (comme signalé

9. La valeur de 1.0 n'a aucune raison de constituer un point de repère essentiel (même si son emploi est équivalent à l'emploi de la méthode de Blondel *et al.*, 2008). Si nous lui accordons un rôle ici, c'est parce qu'il fournit une partition en deux du graphe, intéressante en elle-même. L'expérience tendrait à montrer que les valeurs intéressantes de la résolution se situent souvent entre 0.2 et 3.0.



*supra*) que la détection de communautés cherche à déceler les articulations naturelles du graphe, mais que rien ne permet d’indiquer qu’un découpage obtenu soit plus naturel qu’un autre. La stabilité de ce découpage en deux communautés est peut-être un signe d’une certaine pertinence de ce résultat par rapport à d’autres.



Graphique 1

Cette représentation synthétique permet d’identifier, en deçà de la couche superficielle évoquée plus haut, deux couches de profondeur d’analyse qui semblent *a priori* intéressantes, et correspondant respectivement à des découpages en trois et cinq communautés.

### Un premier niveau de profondeur

Le choix d’une résolution de 1.0 fait apparaître deux communautés de mots, l’une organisée autour d’*excellence* (nœuds blancs), l’autre autour de *projet* (nœuds noirs). Une entrée dans la constitution de chacune des deux communautés donne une idée du parallèle entre la thématique (constitution du dossier de candidature vs idées sur la recherche) et l’orientation de communication (injonction vs argumentation) signalée plus haut.

Chacune des deux communautés dispose d’un mot phare, le plus central<sup>10</sup>, qui apparaît au croisement de quantité de chemins reliant le lexique du texte. Il s’agit respectivement d’*excellence* et de *projet*.

La communauté centrée sur *projet* contient le vocabulaire touchant au document écrit qu’est le dossier de candidature. On trouve à la fois ce qui touche à

10. Au sens de la *betweenness centrality* évoquée dans la note 5.

sa rédaction (*contenu*), sa communication aux instances (*date, heure, adresse, électronique, soumission*, tous mots qui renvoient aux modalités de son envoi électronique) ainsi qu'à sa lecture (*évaluation, jury, examen, issue*).

La communauté centrée sur *excellence* contient, elle, véritablement les enjeux d'idées qui forment l'objet de l'argumentation, et qui renseignent sur la vision que donnent les scripteurs de la recherche et de l'excellence. L'entrée dans cette communauté de mots demande d'en examiner la structure interne, telle qu'elle apparaît à un niveau de profondeur supérieur.

L'observation de cette bipartition du graphe de cooccurrence montre que le contenu du texte est fait de deux champs sémantiques distincts. D'un point de vue méthodologique, l'essentiel est de voir que cette bipartition apparaît spontanément, comme seul résultat du travail de l'algorithme. Cette première étape du travail ne repose pas sur un travail interprétatif fait par le lecteur, qui regrouperait les unités en familles différentes. Ce que montre l'analyse de la modularité faite sur le graphe de cooccurrence, c'est qu'il y a une première couche (en surface) du lexique qui se sépare en deux, du simple fait de la fréquence des relations de cooccurrence. Le travail interprétatif du lecteur humain n'intervient que dans un second temps, pour trouver un sens et une cohérence aux communautés de nœuds identifiées par l'algorithme – si sens et cohérence il y a. En l'occurrence, on peut mettre en évidence les faits évoqués plus haut :

- une première thématique (autour de *projet*) concentrée sur le document, thématique technique dans laquelle aucun ancrage idéologique ne serait décelable ;
- une deuxième thématique (autour d'*excellence*) concentrerait toute la part idéologique et imaginative du contenu du texte.

### *Un deuxième niveau de profondeur : la structure de la communauté centrée sur excellence*

Le choix d'une résolution de 0.8 fait apparaître trois communautés sur le graphe, comme le laisse voir la figure 3 (en annexe). Comme exposé plus haut, ce découpage n'est pas une redistribution des communautés tracées précédemment – et c'est ce fait qui nous permet de proposer une interprétation de ce phénomène sous forme de couches. La communauté centrée sur *excellence* identifiée précédemment se scinde en deux, pour laisser s'autonomiser la petite communauté apparaissant en rouge sur la figure 3. Cette nouvelle communauté est centrée sur le nom *initiative* et permet de réévaluer le contenu de la communauté centrée sur *excellence*.

La communauté nouvellement apparue contient, en plus d'*initiative*, des noms renvoyant à l'organisation de la recherche (*laboratoire, programme*), sa planification (*objectif, engagement, convention*), son financement (*financement, fonds, bénéficiaire, dotation, montant*). Pour comprendre ce qui forme

la cohérence de cette nouvelle communauté, il faut voir que le nom *initiative* appartient en fait à la terminologie du discours en question. Pour ainsi dire, il connaît dans ce texte un emploi figé, puisque la quasi-totalité de ses occurrences dans ce texte (49 sur 55) sont des occurrences d'*initiative d'excellence* – cette expression abrégée par ailleurs en IdEx, nouveau mode de gestion de la recherche que le document en question contribue à inventer. Le nom *excellence* quant à lui ne connaît pas un tel figement (seules 49 occurrences sur 95 sont des *initiatives d'excellence*), comme le montrent notamment ces occurrences (où il est complété par un adjectif, comme *mondiale* ou *universitaire*) :

[...] l'ensemble des activités d'excellence mondiale de recherche et de formation définies par l'Idex et ayant un fort effet de structuration sur le site.  
Excellence universitaire [titre de section]  
[...] démonstration d'une excellence de niveau international dans un ensemble significatif de disciplines, [...]

L'examen du graphe montre que, à ce niveau d'analyse, *excellence* fait encore partie d'une communauté de mots qui ne renvoient pas *a priori* à la technique ou à l'institution, notamment des termes à axiologie positive ou connotés positivement (*ambition, qualité, capacité, avenir*). Là où la communauté centrée sur *initiative* met l'accent sur les moyens (en l'occurrence, financiers), cette communauté est plus tournée vers la finalité et le dynamisme (*dynamique, entraînement, innovation, évolution, insertion*). Le rapport de l'institution au monde extérieur n'est là encore pas seulement financier, mais investit une certaine part du champ économique et politique (*insertion, économique, international, politique*).

C'est-à-dire qu'*initiative*, dans ce texte, renvoie précisément et uniquement à la dimension institutionnelle du texte et de la gestion de l'enseignement supérieur et de la recherche, alors qu'*excellence* garde la possibilité de renvoyer à un niveau de qualité élevé. On peut représenter, dans le tableau 3 *infra*, la réévaluation des contenus dans l'analyse que nécessite le passage d'une couche à l'autre du lexique :

	Autour d'excellence	Autour de projet	
Couche 1 (en surface)	Thématique idéologique, touchant à la gestion et au déroulement de l'enseignement et de la recherche	Thématique de document (écriture et interprétation du projet)	
	Autour d'excellence	Autour d'initiative	Autour de projet
Couche 2 (plus profonde)	Qualités, idéologie et contenus de la recherche et de l'enseignement	Gestion institutionnelle de la recherche et de l'enseignement	Idem

Tableau 3. Affinement de l'analyse des communautés

### Un troisième niveau de profondeur : vers un morcellement ?

Le choix d'une résolution de 0.6 fait apparaître un graphe qui, par rapport aux précédents, montre une topographie bien plus morcelée. En effet, pour sa part, la communauté organisée autour de *projet* n'évolue presque pas en passant du niveau de profondeur précédent à celui-ci, et conserve donc une taille importante (seuls 2 des 26 nœuds qui la composent changent de communauté). En revanche, la communauté organisée originellement autour d'*excellence* – plus importante au départ il est vrai – se morcelle à nouveau, ce qui permet d'envisager sa structure à un plus grand niveau de profondeur.

Couche 1 (en surface)	Autour d'excellence			Autour de <i>projet</i>	
	Thématique idéologique, touchant à la gestion et au déroulement de l'enseignement et de la recherche			Thématique de document (écriture et interprétation du projet)	
Couche 2	Autour d'excellence		Autour d'initiative		Autour de <i>projet</i>
	Qualités, idéologie et contenus de la recherche et de l'enseignement		Gestion institutionnelle de la recherche et de l'enseignement		Idem
Couche 3 (la plus profonde)	Autour de <i>porteur</i>	Autour d'excellence	Autour de <i>recherche</i>	Autour d'initiative	Autour de <i>projet</i>

Tableau 4. Deuxième affinement de l'analyse des communautés

La communauté regroupée autour d'*excellence* concernerait plutôt le fonctionnement de la recherche et ses propriétés, avec un accent mis sur sa dimension politique, ce qui la fait sortir de la sphère des institutions locales (*politique, gouvernance, international, monde*) et l'envisage sous l'angle de sa dynamique (*dynamique, structuration, développement, entraînement, valorisation*).

Elle contient justement des noms faisant référence aux humains (*porteur, partenaire*) ou aux entités abstraites (*cohérence, stratégie*). Ces entités sont, d'une part, intégrées au fonctionnement de la recherche et, d'autre part, impliquées dans le projet, ainsi que l'attestent des constructions comme *porteur de projet, partenaire du projet*.

Autour d'*initiative* se retrouverait seulement la part institutionnelle de la recherche, notamment les moyens financiers (*financement, moyen, investissement, fond, dotation, bénéficiaire*) et les processus dans lesquels les acteurs s'engagent, sans que pour autant on sache réellement ce qu'ils font (*action, engagement, convention*).

La communauté centrée sur *recherche* ferait plutôt référence au terrain effectif sur lequel ont lieu les actions, avec notamment des noms d'activités et des noms collectifs renvoyant à des institutions (*recherche, enseignement, entreprise, activité, établissements, français, étranger, regroupement, institution*).

Une particularité de ce niveau d'analyse est qu'il permet de faire apparaître une communauté qui se situe aux frontières des deux grandes communautés identifiées initialement (au niveau d'analyse le plus superficiel). Il s'agit d'une communauté ne contenant que quatre noms : *cohérence, porteur, partenaire, stratégie*. Un retour à la lettre du texte montre justement que ces noms font référence aux humains (*porteur, partenaire*) ou aux entités abstraites (*cohérence, stratégie*) qui, d'une part, sont intégrés au fonctionnement de la recherche et de l'enseignement et, d'autre part, renvoient aux entités impliquées dans le projet (notamment du fait des emplois *porteur de projet, partenaire du projet*). Ayant des relations avec les deux communautés, ces noms ne peuvent appartenir en fait à aucune d'entre elles, ce qui explique la formation de cette communauté de taille très réduite, et qui n'a aucune autonomie du point de vue du sens. Seul un contraste entre ce niveau de profondeur et un niveau de moindre profondeur permet de mettre en évidence le statut intermédiaire de ces noms.

## **La trajectoire de l'excellence à travers les couches d'analyse en profondeur**

### *L'emploi d'excellence : vue d'ensemble de sa trajectoire*

Étudier la contribution du nom *excellence* au lexique de ce texte suppose de considérer les communautés dont il fait partie à chacune des couches de profondeur dans l'analyse. Le but est d'identifier les termes qui, d'un changement de couche à l'autre, vont rester ses proches voisins, ou au contraire, ne jamais faire partie de ses voisins.

Comme on l'a vu plus haut (section : « Un premier niveau de profondeur »), *excellence* contribue à séparer nettement le champ des idées du texte entre une sphère technique et à court terme (la rédaction et l'examen du dossier de candidature) et une sphère qui concerne les enjeux à long terme, l'organisation et le contenu du travail des institutions (ensemble d'idées auquel *excellence* contribue). C'est-à-dire que, malgré l'appropriation de ce terme par la terminologie officielle, il garde une part de son sens non technique, ambitieux et relevant du domaine des idées.

Par ailleurs, l'examen d'autres niveaux de profondeur montre que, même dans ce discours institutionnel, le nom *excellence* garde des affinités avec le domaine non technique. On assiste en effet peu à peu, au sein de ce qui était la communauté d'*excellence* au niveau minimal de profondeur, à l'autonomisa-

tion de plusieurs champs sémantiques : l'un (au niveau 0.8) renvoyant à la gestion institutionnelle de la recherche, l'autre (au niveau 0.6) concernant les porteurs du projet et la gestion de ces projets au sein de la recherche. C'est dire que, à mesure que l'on creuse dans les couches d'organisation du lexique, *excellence* se désolidarise d'autant plus des contingences à courte durée de la gestion de la recherche pour être lié à des termes renvoyant à des concepts dégagés des obligations matérielles (*ambition, politique, économique, formation*).

Cependant, le passage à de plus grands niveaux de profondeur montre un paysage différent (voir tableau 5 *infra*). En considérant des valeurs très faibles de la résolution, on laisse en effet très peu de temps au marcheur pour parcourir le graphe, et donc on met le doigt sur le réseau de nœuds qui restent le plus étroitement liés au nom de départ. En creusant jusqu'à des niveaux de profondeurs extrêmes, on voit que, en fait, l'attachement du nom *excellence* à la démarche institutionnelle et à la gestion financière reste le plus fort.

Résolution	0.1	0.2	0.3		
Membres de la communauté d'excellence	financement	financement	porteur	financement	innovation
	moyens	modalité	bénéficiaire	moyens	montant
	programme	laboratoire	interface	engagement	stratégie
	convention	programme	montant	ressources	pilotage
	montant	convention	terme	cohérence	crédibilité
		fonds	valorisation	laboratoire	terme
		innovation	institution	programme	contribution
		entraînement	évolution	convention	évolution

Tableau 5. Membres de la communauté d'excellence  
aux niveaux de profondeur maximaux

Interprétation

L'analyse logométrique montrée ici a permis de déceler un traitement particulier du mot *excellence* dans un texte institutionnel. L'organisation lexicale de ce texte reflète au moins en partie une tension entre deux domaines de sens dont le nom *excellence* est le siège (voir « La rhétorique de l'excellence... », *supra*). On a vu en effet que le sens d'*excellence* se partageait en deux valeurs : un sens renvoyant à l'idéal de compétence, et un autre – terme technique – renvoyant à la démarche institutionnelle de gestion de la recherche, en rapport, donc, avec l'évaluation des pratiques de recherche. Cette tension au sein même du texte institutionnel reflète à sa manière – de façon évidemment lacunaire – les tensions, les difficultés et les clivages qu'a connus la notion d'excellence lors

des démarches récentes d'évaluation de la recherche, dont certains discours argumentatifs ou scientifiques tenus par des parties prenantes à ces débats rendent également compte.

Ces derniers discours laissent voir à leur manière un clivage profond entre la chose technique – gestion et évaluation institutionnelle de la recherche – et la dimension idéale, qualitative voire idéologique que lui prêtent ses acteurs. Ceux-ci vivent en effet une tension entre leurs propres idéaux – parmi lesquels « le sens de la vocation est particulièrement fort » (Padis, 2012, p. 15) – et l'évaluation par l'application stricte d'une méthode purement technique (en général, bibliométrique). Une des revendications des chercheurs reste en effet la nécessité de doter l'évaluation d'un sens, basé sur une compréhension de leur travail. Dans les discours, une telle optique est radicalement différente de celle des évaluateurs, dont on a dénoncé le caractère strictement technique, et la distance avec la réalité de la recherche : inadéquation des critères d'évaluation (Terret, Vigarello, 2012), méconnaissance du fonctionnement des disciplines (Bonami, Jaffard, 2012), non-reconnaissance de la diversité au sein d'une discipline (Chiapello, 2012).

La trajectoire que l'on a pu observer (voir « Sur l'emploi d'excellence », *supra*) montre que, à sa manière, le texte étudié ici contribue à la tension au sein de ce domaine d'idées. Le contraste qu'il met en jeu entre la sphère technique et la sphère idéale est un des aspects récurrents de l'évaluation de la recherche telle qu'elle a été conduite.

### *Conclusion méthodologique*

La trajectoire d'*excellence* à travers les différentes couches de lexique montre que les affinités cooccurentielles d'un même terme sont à examiner à différents niveaux de profondeur. Chacun de ces niveaux correspond à un regard plus ou moins fin sur les affinités de ce mot avec le reste du lexique. L'analyse faite ici montre que l'on peut constater des allers et retours dans l'interprétation, et que l'affinité entre un mot et une thématique peut n'être effective qu'à un niveau d'analyse et se perdre à un autre niveau. Cela ne veut pas dire pour autant que les niveaux les plus profonds soient les plus significatifs. Le plus significatif pourrait résider dans la dynamique de l'interprétation. C'est voir que, à un niveau de détail très fin, *excellence* reste proche de questions financières (ce que l'on peut sans doute attester par un simple relevé de cooccurrences ponctuelles) et que, concernant l'économie du texte dans sa globalité, il contribue au contraire aux thématiques touchant aux idéologies, à la planification du travail, et cela dans une perspective à long terme. C'est donc voir que, malgré les plus proches voisins cooccurentiels qui le rattachent à une sphère technique, la position qu'il prend dans l'ensemble de la cooccurrence généralisée le dégage de cette sphère.

La méthode exposée ici permet donc de mettre en évidence une autre organisation des relations de cooccurrence dans un texte. Les études de la cooccurrence généralisée ont, depuis longtemps, mis en avant le caractère modulaire du lexique d'un texte, en le répartissant en familles. Nous proposons ici de voir que cette cooccurrence généralisée est dotée par ailleurs d'une géométrie « en profondeur », reflétant les plus ou moins grandes affinités cooccurrenceuses à différents niveaux d'analyse.

## Références

- BLONDEL Vincent, 2011, «The Louvain method for community detection in large networks» [<http://perso.uclouvain.be/vincent.blondel/research/louvain.html>], consulté le 8 mars 2013.
- BLONDEL Vincent, GUILLAUME Jean-Loup, LAMBIOTTE Renaud, LEFEBVRE Étienne, 2008, «Fast unfolding of communities in large network», *Journal of Statistical Mechanics. Theory and Experiment*.
- BONAMI Aline, JAFFARD Stéphane, 2012. «Les mathématiques à l'épreuve de "l'excellence"», *Esprit*, n° 386, p. 30-40.
- CHEN Zheng, JI Heng, 2010, «Graph-based clustering for computational linguistics : a survey», *Proceedings of the 2010 Workshop on Graph-based Methods for Natural Language Processing*, ACL 2010, p. 1-9.
- CHIAPELLO Ève, 2012, «Écoles de commerce. La pression de l'internationalisation», *Esprit*, n° 386, p. 18-25.
- CUSSÓ Roser, GOBIN Corinne, 2008, «Du discours politique au discours expert. Le changement politique mis hors débat?» *Mots. Les langages du politique*, n° 88, p. 5-11.
- GOBIN Corinne, DEROUBAIX Jean-Claude, 2010, «L'analyse du discours des organisations internationales. Un vaste champ encore peu exploré», *Mots. Les langages du politique*, n° 94, p. 107-114.
- GRANGET Lucia, 2009, «Les universités en quête de prestige dans le grand jeu de la concurrence», *Communication et organisation*, n° 35, p. 148-157.
- LAMBIOTTE Renaud, DELVENNE Jean-Charles, BARAHONA Mathias, 2009, «Laplacian dynamics and multiscale modular structure in networks» [<http://arxiv.org/pdf/0812.1770v3.pdf>], consulté le 8 mars 2013.
- LUONG Xuan, BRUNET Étienne, LONGRÉE Dominique, MAYAFFRE Damon, MELLET Sylvie, POUDAT Céline, 2010, «La cooccurrence, une relation asymétrique?» *JADT* 2010, p. 321-331.
- NEWMAN Mark E. J., 2006, «Modularity and community structure in networks», *Proceedings of the National Academy of Science*, n° 103 (23), p. 8577-8582.
- PADIS Marc-Olivier, 2012, «Quelle évaluation? Quelle excellence?» *Esprit*, n° 386, p. 13-17.
- RÉMI-GIRAUD Sylvianne, 2010, «Sémantique lexicale et langages du politique. Le paradoxe d'un mariage difficile?» *Mots. Les langages du politique*, n° 94, p. 165-173.



- TAUVERON Matthias, 2012, « Incrémentation lexicale dans les textes : une auto-organisation », *Actes de la conférence conjointe JEP-TALN-RECITAL 2012*, vol. III : *RECITAL*, p. 15-28.
- TERRET Thierry, VIGARELLO Georges, 2012, « Sciences humaines et sciences exactes. Le déséquilibre de l'évaluation », *Esprit*, n° 386, p. 26-29.
- VÉRONIS Jean, 2004, « HyperLex. Lexical cartography for information retrieval », *Computer Speech & Language*, n° 18 (3), p. 223-252.
- VIPREY Jean-Marie, 2006, « Structure non séquentielle des textes », *Langages*, n° 163, p. 71-85.

Annexe



Figure 1. Représentations du lexique d'un même texte à trois niveaux de profondeur\* différents (respectivement 1.2, 0.8 et 0.6)

\*Pour une explication de cette métaphore, voir la section : « Proposition d'analyse : la superposition des couches de noeuds », *supra*.

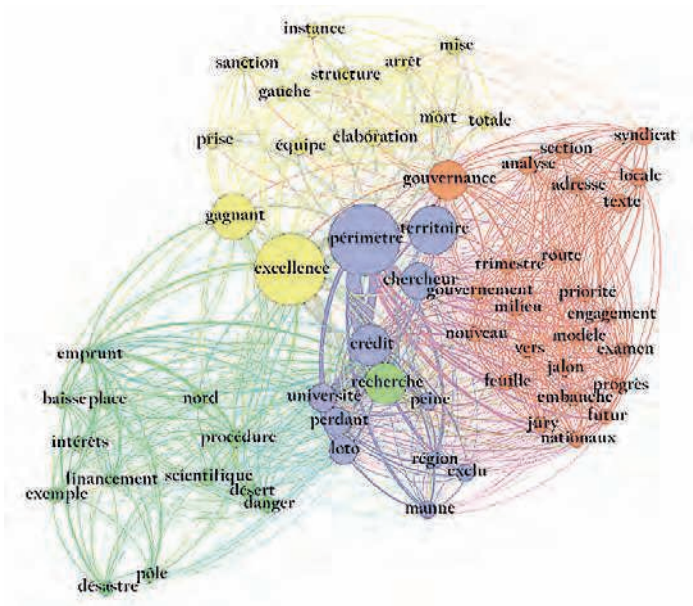


Figure 2. Exemple de graphe de cooccurrence

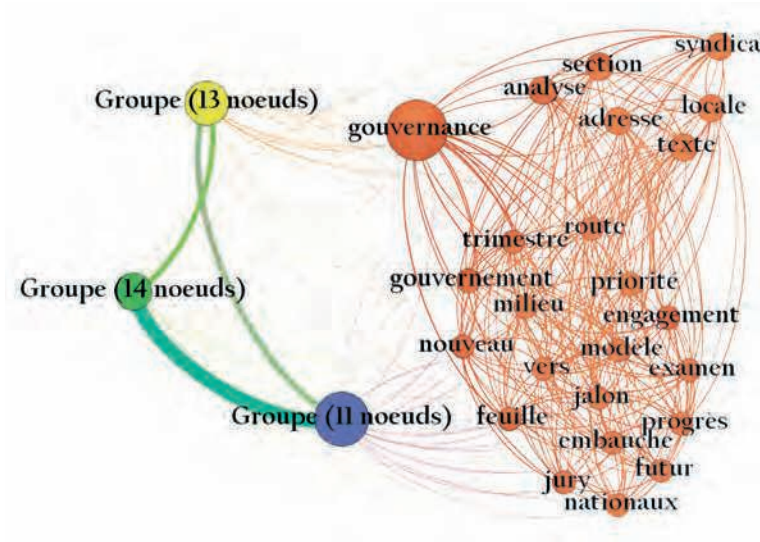


Figure 3. Graphe de cooccurrence pour le texte AAPIE. Représentation de la modularité avec une résolution de 0.8